

Title: Technical note: Analysis of observation uncertainty for flood assimilation and forecasting

Manuscript ID: hess-2018-43

Authors: J. A. Waller, J. García-Pintado, D. C. Mason, S. L. Dance, N. K. Nichols

We thank the reviewer for their positive comments, which will help improve the manuscript. Below we give each comment in bold (abridged where appropriate) and describe how we plan to alter the manuscript to address the reviewer's concern. We give suggested changes to the manuscript in italic font. We note that several of the comments ask for clarification/additional details about the derivation of the water level observations and their associated uncertainties. To address these comments we would replace Section 3 by a methodology section with the following subsections:

3.1 Derivation of water level observations.

3.2 Model and data assimilation.

3.3 Quality control and data thinning.

3.4 Potential observation error sources.

3.5 Experimental design.

The section would contain material from the current Section 3, some of the material that is currently in the introduction and include some new additional material that describes in more detail the observation derivation process and the possible sources of observation uncertainty.

Response

1. **Title: I wonder whether flood assimilation is the most appropriate wording. The authors might consider a slight change in the title (assimilation in flood forecast models).**

After considering this and other comments made by the reviewers we propose a new title for the manuscript: *Technical note: Assessment of observation quality for data assimilation in flood models*

2. **Page 1, Lines 7-8 The estimated correlations do not behave as expected: could you please be more explicit?.**

We would alter the abstract as follows so that it is more explicit:

'The average estimated correlation length scale of 7km is longer than the expected value of 250m. Furthermore, the correlations do not decrease monotonically; this unexpected behaviour is shown to be the result of assimilating some anomalous observations.'

3. **Page 2, line 13 unbiased: is bias the only error type in Remote Sensing-derived observations of floods?**

Bias is not the only type of error, but it is assumed that any systematic error is removed from an observation before assimilation. We would clarify this in Section 2: *'The observation-minus-background residuals $\mathbf{d}_b^o = \mathbf{y} - \mathcal{H}(\mathbf{x}^b)$, also known as the innovations, are assumed to be unbiased. Hence any bias should be removed before assimilation [Dee, 2005].'*

To avoid confusion we would remove *'if they are to be unbiased'* from Page 2, line 13.

4. **Page 2, line 18 as is standard in other assimilation applications: I suggest rephrasing this sentence. Could you please provide examples of the other assimilation applications?**

To improve fluency and give examples we would reword follows: *‘Thus, to reduce the number of correlated observations and to avoid dealing with spatial correlation in the assimilation, the current approach is to further thin the data (as is standard in other assimilation applications such as NWP and oceanography [Dando et al., 2007, Li et al., 2010]).’*

5. **Page 2, line 19 typically retaining approximately 1% of the pre-thinned observations: I suggest adding at least a reference.**

We would add Mason et al. [2012] as the reference.

6. **Page 2, line 20 The authors might consider commenting on the algorithm presented Mason et al. [2012].**

We would add additional information on the algorithm presented in Mason et al. [2012] in the proposed new sections.

In Section 3.1 we would provide a brief description of water level observation derivation. This would include how:

- flood extent can be extracted from SAR images by determining regions of low backscatter;
- the initial flood extent estimate can be refined;
- the flood extent is intersected with the DEM to provide the WLOs.

In Section 3.2 we would discuss the model and data assimilation. This section would consist of some current material (pg 6 lines 1 - 5) and some new material that describes how the model can be compared to the observations.

In section 3.3 we would provide a brief discussion on the quality control and data thinning. This section would consist mainly of current material (e.g pg 2 lines 12 - 20), but would have some additional information about the quality control and thinning procedures. For example we would include that

‘An additional background check is performed where observations that result in anomalous observation-minus-background residuals are discarded.’

and

‘The applied thinning, as described in Mason et al. [2012], uses a top down clustering approach in which principal component analysis is used to select observations that have the highest information content. The spatial autocorrelation of the resulting observations is calculated, and if any significant correlation exists the thinning procedure is applied iteratively until no significant correlation remains.’

7. **Page 2, lines 21-24: is the discussion on DA of satellite-derived soil moisture strictly relevant here?**

We would remove this sentence.

8. **Page 2, lines 24-30: I believe that moving this paragraph after line 11 could improve the readability of the manuscript. I suggest discussing error types, data thinning, and uncertainty estimation after this general introduction.**

The discussion of error types and data thinning would be moved to the proposed Sections 3.4 and 3.3 respectively (see responses to comment 6 and Reviewer 2 comment 7). As a result the paragraph in lines 24-30 would be moved to after line 11.

9. Page 2, line 31 directly: is this the most appropriate word? Do the authors mean computed in a systematic way?

We would rephrase the sentence as: *'The true state of the system is not known and therefore the exact observation errors and their uncertainties cannot be calculated. Instead observation uncertainties must be estimated statistically [e.g. Hollingsworth and Lönnberg, 1986, Ueno and Nakamura, 2016].'*

10. Page 2, line 34 with good results: would it be possible to clarify this statement?

Rather than 'with good results' we would state *'The diagnostic has been applied to operational numerical weather prediction (NWP) settings to estimate observation uncertainties [Stewart et al., 2014, Waller et al., 2016a,b, Cordoba et al., 2017]. The use of these estimated statistics in NWP results in a more accurate analysis and improvements in objectively measured forecast skill [Weston et al., 2014, Bormann et al., 2016, Campbell et al., 2017].'*

11. Page 3, line 4: is a new paragraph required?

We would join the text to the previous paragraph.

12. Page 3, line 5-6 Thus, we then consider: could you please rephrase this sentence?(e.g. Consequently, ...).

We would rephrase this sentence to state *'To determine the cause of these anomalous results we consider if observations in different sub-domains have different error characteristics. We also consider if the error statistics differ for different phases of the flood event.'*

13. Page 3, line 7 is related not related: please correct this.

We would remove the first 'related'

14. Page 3, lines 8-9 we show that: how is this result related to the papers cited in page 2 line 34? I would like to recommend adding a sentence to clarify the novelty of the study presented in the manuscript.

We would make it clear in the introduction that *'To the best of our knowledge this is the first time that the diagnostics have been applied to estimate error statistics for hydrological data assimilation.'*

15. Page 3, lines 11-12: could you please improve the fluency of this sentence?

We would rephrase as *'Data assimilation is a technique used to provide the best estimate, the analysis, of the current state of a dynamical system. The analysis is denoted $\mathbf{x}^a \in \mathbb{R}^{N^m}$.'*

16. Page 3, line 15: are the superscripts correct?

We would swap the superscripts so the manuscript reads $\mathcal{H} : \mathbb{R}^{N^m} \rightarrow \mathbb{R}^{N^p}$.

17. Page 3, line 21: could you please rephrase the last part of the sentence?

We would rephrase the last part of the sentence as *'H is defined as the observation operator linearised about the background state.'*

18. Page 4, line 30 error is repeated.

We would remove the repeated ‘error’.

19. Page 4, line 30 gross error measurement: could please the authors clarify this statement?

We would rephrase as follows: *‘Data assimilation techniques can lose accuracy if presented with an observation that is grossly inconsistent with the model state [Vanden-Eijnden and Weare, 2013].’*

20. Page 5 line 2 typically: what do the authors mean here? Typically in this dataset or typically in the literature?

We would clarify that *‘Typically the dataset in this study is thinned to a separation distance of 250m.’*

21. Page 5, figure 1: the authors might consider adding the underlying map (or at least the river network).

We feel that the figure may be too cluttered if the observations are plotted over an underlying map. Therefore we propose to add an additional figure panel alongside the current figure (as Figure 1a) showing the digital elevation model over the domain.

22. Page 6, lines 9-10: could the authors please clarify this statement?

We would clarify this comment as follows: *‘As we use an LETKF we must use a modified form of the diagnostic (see Section 2). As a result we are not able to calculate observation error correlations for observation pairs with a separation distance greater than 19km’*

23. Page 6, lines 17-18: please correct capital letters, full stop.

We would correct the capital letters and full stop.

24. Page 6, line 20 than those we estimate: I suggest rephrasing this sentence.

We would rephrase to *‘Therefore, we would expect the true standard deviations and length scales to be larger than those we estimate using the diagnostic.’*

25. Page 6, line 31: the authors might consider replacing shoulder with something more formal.

We would replace any occurrences of the word ‘shoulder’ with ‘local maximum’.

26. Page 7, line 6: is are resulting in the increase... correct? I wonder whether the authors mean that observations in different areas lead to the increase of...

We would replace the sentence with *‘It is possible that the local maximum in the correlations is a result of observations on different tributaries of the river.’*

27. Page 7, line 10 and therefore the results are subject to greater sampling error: could the authors please clarify this statement?

We would clarify this statement as follows: *‘We note that there are fewer observations in the eastern domain. This results in fewer available samples for the calculation in equation (3) and hence the results are subject to greater sampling error.’*

28. Page 8, line 6 we see: I suggest replacing this with something more formal.

We would start the sentence with *‘At the beginning of the flood period’*.

29. Page 8, line 12: please rephrase this sentence.

We would rephrase as follows: ‘*Fig.7 shows the estimated error statistics for the recession stages of the flood. During this period a high proportion of the observations were in areas which remained flooded but were disconnected from the main river flow.*’

30. Page 8, line 16 to keep the forecast on track: could the authors please clarify this statement?

We would rephrase as follows: ‘*the assimilation increments were of a smaller magnitude (thus, not so effective) in these last stages.*’

31. Page 9, line 6: the authors might consider replacing end of the flood with receding limb or something more formal than end.

We would replace ‘end of the flood’ with ‘during the receding limb of the flood’.

32. Page 9, line 9 More study is needed in this context: the authors might want to add details on future work/research needs.

We would alter the sentence so it states: ‘*However, due to the dependence of the observation error on the choice of observation operator and model resolution, results will differ for each individual user. Therefore, further study may be required to understand how the diagnostic results can best support QC protocols.*’

References

- N. Bormann, M. Bonavita, R. Dragani, R. Eresmaa, M. Matricardi, and A. McNally. Enhancing the impact of iasi observations through an updated observation-error covariance matrix. *Quarterly Journal of the Royal Meteorological Society*, 142(697):1767–1780, 2016. doi: 10.1002/qj.2774.
- W. F. Campbell, E. A. Satterfield, B. Ruston, and N. L. Baker. Accounting for correlated observation error in a dual-formulation 4d variational data assimilation system. *Monthly Weather Review*, 145(3): 1019–1032, 2017. doi: 10.1175/MWR-D-16-0240.1.
- M. Cordoba, S. Dance, G. Kelly, N. Nichols, and J. Waller. Diagnosing atmospheric motion vector observation errors for an operational high resolution data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 143(702):333–341, 2017. doi: 10.1002/qj.2925.
- M. Dando, A. Thorpe, and J. Eyre. The optimal density of atmospheric sounder observations in the met office NWP system. *Quarterly Journal of the Royal Meteorological Society*, 133:1933–1943, 2007.
- D. P. Dee. Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613): 3323–3343, 2005. doi: 10.1256/qj.05.137.
- A. Hollingsworth and P. Lönnberg. The statistical structure of short-range forecast errors as determined from radiosonde data. part i: The wind field. *Tellus A*, 38A(2):111–136, 1986. doi: 10.1111/j.1600-0870.1986.tb00460.x.
- X. Li, J. Zhu, Y. Xiao, and R. Wang. A model-based observation-thinning scheme for the assimilation of high-resolution sst in the shelf and coastal seas around china. *Journal of Atmospheric and Oceanic Technology*, 27(6):1044–1058, 2010. doi: 10.1175/2010JTECHO709.1.

- D. Mason, G.-P. Schumann, J. Neal, J. Garcia-Pintado, and P. Bates. Automatic near real-time selection of flood water levels from high resolution synthetic aperture radar images for assimilation into hydraulic models: A case study. *Remote Sensing of Environment*, 124(Supplement C):705 – 716, 2012. doi: <https://doi.org/10.1016/j.rse.2012.06.017>.
- L. M. Stewart, S. L. Dance, N. K. Nichols, J. R. Eyre, and J. Cameron. Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. *Quarterly Journal of the Royal Meteorological Society*, 140:1236–1244, 2014. doi: 10.1002/qj.2211.
- G. Ueno and N. Nakamura. Bayesian estimation of the observation-error covariance matrix in ensemble-based filters. *Quarterly Journal of the Royal Meteorological Society*, 142(698):2055–2080, 2016. doi: 10.1002/qj.2803.
- E. Vanden-Eijnden and J. Weare. Data assimilation in the low noise regime with application to the kuroshio. *Monthly Weather Review*, 141(6):1822–1841, 2013. doi: 10.1175/MWR-D-12-00060.1.
- J. A. Waller, S. P. Ballard, S. L. Dance, G. Kelly, N. K. Nichols, and D. Simonin. Diagnosing horizontal and inter-channel observation error correlations for SEVIRI observations using observation-minus-background and observation-minus-analysis statistics. *Remote Sensing*, 8(7):581, 2016a. doi: 10.3390/rs8070581.
- J. A. Waller, D. Simonin, S. L. Dance, N. K. Nichols, and S. P. Ballard. Diagnosing observation error correlations for Doppler radar radial winds in the Met Office UKV model using observation-minus-background and observation-minus-analysis statistics. *Monthly Weather Review*, 144(10):3533–3551, 2016b. doi: 10.1175/MWR-D-15-0340.1.
- P. P. Weston, W. Bell, and J. R. Eyre. Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society*, 140(685):2420–2429, 2014. doi: 10.1002/qj.2306.